



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : <b>C12N 15/54, 9/12, C07K 14/415, C12N 5/10, G01N 33/50, C12Q 1/68, 1/48</b>		A1	(11) International Publication Number: <b>WO 00/36117</b>
			(43) International Publication Date: 22 June 2000 (22.06.00)
(21) International Application Number: PCT/US99/29826			(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 15 December 1999 (15.12.99)			
(30) Priority Data: 60/112,558 16 December 1998 (16.12.98) US			
(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). FALCO, Saverio, Carl [US/US]; 1902 Miller Road, Arden, DE 19810 (US). KINNEY, Anthony, J. [GB/US]; 609 Lore Avenue, Wilmington, DE 19809 (US).			
(74) Agent: FEULNER, Gregory, J.; E.I. du Pont de Nemours and Company, Legal Patent Center, 1007 Market Street, Wilmington, DE 19898 (US).			Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PHOSPHOLIPID BIOSYNTHETIC ENZYMES

```

** ***** ** * * * * *
SEQ ID NO:14 M-----P--SVLYIPNIIGYFRIIINFIAFVCYSNKALFAILYFISFVLGDGVDGW
SEQ ID NO:16 MAQPSSKKTP--SVLYIPNIIGYFRIIINFIAFVCYSNRVLFAILYFFSFCDGLDGW
SEQ ID NO:20 MAKKPGPRSSKLSVLYIPNIIGYIRVLLNCFASFQCLSNKILFSILYFLSFVCDVADGW
SEQ ID NO:21 MAKKERPRPEKLSVLYIPNIVGYMRVLLNCVAFVCFSNKPLFSVLYFFSFCCDAVDGW
1 60

** * * * *
SEQ ID NO:14 FARKFNQASTFGAVLDMVTDVSTACLLALLSQFYRPGVLVFLILLGLDITSHWFQMYSSF
SEQ ID NO:16 FARXFNQASTCGAVLDMVTDVSTACLLALLSQFYRPGVLVFLILLGLDITSHWFQMYSSF
SEQ ID NO:20 CARXFNQVSTFGAVLDMVTDVSTACLLVVLSQLYKPLSFLSLLALDIASHWLQMYSTF
SEQ ID NO:21 VARXFNQVSTFGAVLDMVTDVSTACLLVILSQIYRPSLVFLSLLALDIASHWLQMYSTF
61 120

* * * * *
SEQ ID NO:14 LSGKTSBKDVKHTGNWLLKLYGYRPFMAFCCVSCVLYIFLFLFADEESTSLIVCKGI
SEQ ID NO:16 LSGKTSBKDVKHTGNWLLKLYGYRPFMAFCCVSCVLYIFLFLFADEESTSLIVCKGI
SEQ ID NO:20 LTGKTSBKDVKHTGNWLLKLYGYRPFMAFCCVSCVLYIFLFLFADEESTSLIVCKGI
SEQ ID NO:21 LAGKSSBKDVKHTGNWLLKLYGYRPFMAFCCVSCVLYIFLFLFADEESTSLIVCKGI
121 180

* * * * *
SEQ ID NO:14 LNQSPVVLVSTLVGAVKQATNVIQMKTAADACVVDLKRKSK---
SEQ ID NO:16 LKQSPVVLVSTLVGAVKQATNVIQMKTAADACVVDLKRKSK---
SEQ ID NO:20 LQKISPLSLLMGTSLFGWAVKQATNVIQMKTAADACVVDLKRKSK---
SEQ ID NO:21 LTQISPLSLLMGTSLFGWAVKQATNVIQMKTAADACVVDLKRKSK---
181 228
```

## (57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a phospholipid biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the phospholipid biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the phospholipid biosynthetic enzyme in a transformed host cell.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## TITLE

### PHOSPHOLIPID BIOSYNTHETIC ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/112,558, filed December 16, 1998.

5

## FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding phospholipid biosynthetic enzymes in plants and seeds.

## BACKGROUND OF THE INVENTION

10 CDP-alcohol phosphatidyltransferase is an enzyme able to catalyze the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and breaking of a phosphoride anhydride bond. An example of a CDP-alcohol phosphatidyltransferase is phosphatidylglycerophosphate synthase, also called CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5) which is  
15 an integral membrane protein that catalyzes the committed step in the synthesis of acidic phospholipids, the formation of phosphatidylglycerophosphate from CDP-diacylglycerol (phosphatidyl-CMP) and glycerol-3-phosphate (Carman and Belunis (1983) *Can J Microbiol* 29:1452-1457). In contrast to the other enzymes involved in the diphosphatidylglycerol biosynthetic pathway, phosphatidyl-glycerophosphate synthase is highly regulated, both by  
20 cross-pathway control and by factors affecting mitochondrial development (Minskoff, S. A. and Greenberg, M. L. (1997) *Biochim Biophys Acta* 1348:187-191). This enzyme has been extensively studied in bacteria and yeasts (*S. cerevisiae* and *S. pombe*). In *S. cerevisiae*, the PGS1 gene encodes phosphatidylglycerophosphate synthase (Chang et al. (1998) *J Biol Chem* 273:9829-9836). Disruption of the PGS1 gene does not lead to unconditional lethality  
25 but does result to dependence on a fermentable carbon source for growth, temperature sensitivity of growth, and a petite lethal phenotype (Chang et al. (1998) *J Biol Chem* 273:9829-9836), demonstrating the importance of phosphatidylglycerophosphate synthase for mitochondrial function.

Significantly less is known about phosphatidylglycerophosphate synthase and its  
30 function in plants. Phosphatidylglycerophosphate is the only phospholipid in chloroplast thylakoids such that by manipulating phosphatidylglycerophosphate synthase expression it may be possible to engineer cold resistance.

Phosphatidylinositol synthase, also called CDP-diacylglycerol:myo-inositol 3-phosphatidyl-transferase (EC 2.7.8.11) is a membrane-bound enzyme present in all  
35 mammalian cells and localized predominantly to the endoplasmic reticulum. Phosphatidylinositol synthase catalyzes the last step in the phosphatidylinositol biosynthesis. This step involves the condensation of CDP-diacylglycerol and myo-inositol to form phosphatidylinositol and CMP. Besides being an essential membrane

phospholipid, phosphatidylinositol is involved in protein membrane anchoring and is the precursor of inositol-tri-phosphate and diacylglycerol (Antonsson, B. (1997) *Biochim Biophys Acta* 1348:179-186). Phosphatidylinositol synthase has been purified from yeast microsomes and its activity shown to be dependent on the presence of  $Mn^{+2}$  or  $Mg^{+2}$  and detergents. The expression of phosphatidylinositol synthase is insensitive to the addition of myo-inositol and choline to culture medium or the transition of growth phase. Disruption of PIS, the gene encoding phosphatidylinositol synthase in *S. cerevisiae* is lethal suggesting that the function of the phosphatidylinositol synthase gene is essential for progression of the yeast cell cycle (Nikawa, J. and Yamashita, S. (1997) *Biochim Biophys Acta* 1348:173-178).

### SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 50 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a soybean CDP-alcohol phosphatidyltransferase polypeptide of SEQ ID NO:2 and a wheat CDP-alcohol phosphatidyltransferase polypeptide of SEQ ID NO:6. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a soybean CDP-alcohol phosphatidyltransferase polypeptide of SEQ ID NO:4. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a soybean phosphatidylglycerophosphate synthase polypeptide of SEQ ID NO:10, and a soybean phosphatidylglycerophosphate synthase polypeptide of SEQ ID NO:12. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 150 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a corn phosphatidylglycerophosphate synthase polypeptide of SEQ ID NO:8. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 50 amino acids that has at least 90% identity based on the Clustal method of alignment when compared to a soybean phosphatidylinositol synthase polypeptide of SEQ ID NO:18. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a corn phosphatidylinositol synthase polypeptide of SEQ ID NO:14. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 150 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a soybean

phosphatidylinositol synthase polypeptide of SEQ ID NO:20. The present invention also relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 200 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a rice phosphatidylinositol synthase polypeptide of SEQ ID NO:16. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

It is preferred that the isolated polynucleotides of the claimed invention consist of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20. The present invention also relates to an isolated polynucleotide comprising a nucleotide sequences of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 and the complement of such nucleotide sequences.

The present invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to suitable regulatory sequences.

The present invention relates to an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell. The present invention also relates to a virus, preferably a baculovirus, comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

The present invention relates to a process for producing an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting an isolated compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

The present invention relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 50 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2 and 6. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 100 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide of SEQ ID NO:4. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 100 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:10 and 12. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or

phosphatidylinositol synthase polypeptide of at least 150 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide of SEQ ID NO:8. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 50 amino acids comprising at least 90% homology based on the Clustal method of alignment compared to a polypeptide of SEQ ID NO:18. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 100 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide of SEQ ID NO:14. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 150 amino acids comprising at least 85% homology based on the Clustal method of alignment compared to a polypeptide of SEQ ID NO:20. The present invention also relates to a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide of at least 200 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NO:16.

The present invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide in a host cell, preferably a plant cell, the method comprising the steps of: (a) constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; (b) introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; (c) measuring the level of a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide in the host cell containing the isolated polynucleotide; and (d) comparing the level of a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide in the host cell containing the isolated polynucleotide with the level of a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide in the host cell that does not contain the isolated polynucleotide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide gene, preferably a plant CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polypeptide gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a

nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a CDP-  
5 alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase amino acid sequence.

The present invention also relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol  
10 synthase polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

A further embodiment of the instant invention is a method for evaluating at least one  
15 compound for its ability to inhibit the activity of a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase, operably linked to suitable  
20 regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase in the transformed host cell; (c) optionally purifying the CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase  
25 or phosphatidylinositol synthase expressed by the transformed host cell; (d) treating the CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase with a compound to be tested; and (e) comparing the activity of the CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase that has been treated with a test compound to the activity of an  
30 untreated CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase, thereby selecting compounds with potential for inhibitory activity.

The present invention relates to a composition, such as a hybridization mixture, comprising an isolated polynucleotide of the present invention.

35 The present invention relates to an isolated polynucleotide of the present invention comprising at least one of 30 contiguous nucleotides derived from a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19.

The present invention relates to an expression cassette comprising an isolated polynucleotide of the present invention operably linked to a promoter.

The present invention relates to a method for positive selection of a transformed cell comprising: (a) transforming a host cell with the chimeric gene of the present invention or  
 5 an expression cassette of the present invention; and (b) growing the transformed host cell, preferably plant cell, such as a monocot or a dicot, under conditions which allow expression of the CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase polynucleotide in an amount sufficient to complement a yeast *pis* or *pgs1* mutation to provide a positive selection means.

10

### BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 depicts the amino acid alignment between the phosphatidylinositol synthase  
 15 encoded by the nucleotide sequences derived from corn clone cbn10.pk0005.e5 (SEQ ID NO:14), rice clone rl0n.pk0062.f12 (SEQ ID NO:16), soybean clone sl2.pk129.112 (SEQ ID NO:20), and a phosphatidylinositol synthase-encoding nucleic acid fragment from *Arabidopsis thaliana* (NCBI GenBank Identifier (GI) No. 3367632) (SEQ ID NO:21). Amino acids which are conserved among all and at least two sequences with an amino acid  
 20 at that position are indicated with an asterisk (\*) above them. Dashes are used by the program to maximize alignment of the sequences.

Figure 2 depicts the amino acid alignment between the CDP-alcohol  
 phosphatidyltransferase encoded by the nucleotide sequences derived from a contig  
 assembled from a portion of the cDNA insert in soybean clone sdp3c.pk005.h18 and the  
 25 entire cDNA insert in soybean clone sfl1.pk0072.c4 (SEQ ID NO:4) and a CDP-alcohol  
 phosphatidyltransferase-encoding nucleic acid fragment from *Arabidopsis thaliana* (NCBI  
 GenBank Identifier (GI) No. 4115948 (SEQ ID NO:22). Amino acids which are conserved  
 between the two sequences are indicated with an asterisk (\*) above them. Dashes are used  
 by the program to maximize alignment of the sequences.

30 Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. Table 1 also identifies the cDNA clones as individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the  
 35 indicated cDNA clones ("FIS"), contigs assembled from two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs ("Contig\*"), or sequences encoding the entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"). Nucleotide SEQ ID NOs:1, 5, 7, 9, 13, 15, and 17 correspond to nucleotide SEQ ID NOs:11, 13, 1,

- 3, 5, 7, and 9, respectively, presented in U.S. Provisional Application No. 60/112,558, filed December 16, 1998. Amino acid SEQ ID NOs:2, 6, 10, 14, 16, and 18 correspond to amino acid SEQ ID NOs:12, 14, 4, 6, 8, and 10, respectively, presented in U.S. Provisional Application No. 60/112558, filed December 16, 1998. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

**TABLE 1****Phospholipid Biosynthetic Enzymes**

Protein	Clone Designation	Status	SEQ ID NO:	
			(Nucleotide)	(Amino Acid)
CDP-Alcohol Phosphatidyltransferase (Soybean)	sfl1.pk0072.c4	FIS	1	2
CDP-Alcohol Phosphatidyltransferase (Soybean)	Contig of: sdp3c.pk005.h18 sfl1.pk0072.c4 (FIS)	CGS	3	4
CDP-Alcohol Phosphatidyltransferase (Wheat)	wre1n.pk191.g3	FIS	5	6
Phosphatidylglycero-phosphate Synthase (Corn)	Contig of: p0005.cbmev33r cs1.pk0038.g4 (FIS)	Contig*	7	8
Phosphatidylglycero-phosphate Synthase (Soybean)	Contig of: sfl1.pk0086.d7 (FIS) sdp2c.pk013.h24	Contig*	9	10
Phosphatidylglycero-phosphate Synthase (Soybean)	Contig of: sfl1.pk0086.d7 (FIS) sdp2c.pk013.h24 scr1c.pk004.c14	Contig*	11	12
Phosphatidylinositol Synthase (Corn)	cbn10.pk0005.e5 (FIS)	CGS	13	14
Phosphatidylinositol Synthase (Rice)	rl0n.pk0062.f12 (FIS)	CGS	15	16
Phosphatidylinositol Synthase (Soybean)	Contig of: sl2.pk129.l12 src3c.pk022.d15	Contig	17	18
Phosphatidylinositol Synthase (Soybean)	sl2.pk129.l12 (FIS)	CGS	19	20

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.*

219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

#### DETAILED DESCRIPTION OF THE INVENTION

5 In the context of this disclosure, a number of terms shall be utilized. As used herein, a “polynucleotide” is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA,  
10 synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least one of 60 contiguous nucleotides, preferably at least one of 40 contiguous nucleotides, most preferably one of at least 30 contiguous nucleotides derived from SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, or the complement of such sequences.

As used herein, “contig” refers to a nucleotide sequence that is assembled from two or  
15 more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be  
20 assembled into a single contiguous nucleotide sequence.

As used herein, “substantially similar” refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. “Substantially similar” also refers to nucleic acid fragments wherein changes in  
25 one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. “Substantially similar” also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting  
30 transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid  
35 fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment

representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19 and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of a polypeptide (CDP-alcohol phosphatidyl-transferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase) in a host cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial, or viral) may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; measuring the level of a polypeptide in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide in the host cell containing the isolated polynucleotide with the level of a polypeptide in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least about 50 amino acids, preferably at least about 100 amino acids, more preferably at least about 150 amino acids, still more preferably at least about 200 amino acids, and most preferably at least about 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein

or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly,

the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature.

Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

“Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a nucleotide sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be

found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

5       The “translation leader sequence” refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have  
10       been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

      The “3' non-coding sequences” refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid  
15       tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

      “RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA  
20       sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into polypeptide by the cell. “cDNA” refers to a double-stranded DNA that is complementary to and derived from mRNA. “Sense” RNA refers to an RNA transcript that includes the mRNA and so can be translated into a  
25       polypeptide by the cell. “Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or  
30       the coding sequence. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

      The term “operably linked” refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable  
35       of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed.

"Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several phospholipid biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated

by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably one of at least 40, most preferably one of at least 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide. The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a polypeptide of a gene (such as CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase) preferably a substantial portion of a plant polypeptide of a gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19 and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a polypeptide (CDP-alcohol phosphatidyltransferase, phosphatidylglycerophosphate synthase or phosphatidylinositol synthase).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of glycerophospholipids in those cells. Phosphatidylglycerophosphate synthase may be useful for engineering cold resistance since phosphatidylglycerophosphate is the only phospholipid in chloroplast thylakoids. The polypeptides disclosed herein may also be used as targets for herbicide discovery and design since they catalyze various crucial steps in phospholipid metabolism. Their importance to organism viability is demonstrated by tight regulation of phosphatidylglycerophosphate synthase in yeast (Minskoff and Greenberg (1997) *Biochim Biophys Acta* 1348:187-191) and lethality of phosphatidylinositol synthase gene disruption in *S. cerevisiae* (Nikawa and Yamashita (1997) *Biochim Biophys Acta* 1348:173-178).

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter  
5 sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host  
10 plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and  
15 thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different  
20 cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) with or without removing targeting sequences that are already present.  
25 While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the  
30 instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in  
35 reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds, and is not an inherent part of the invention. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded phospholipid biosynthetic enzyme. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 8).

Additionally, the instant polypeptides can be used as a targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the polypeptides described herein catalyze various steps in phospholipid metabolism. Accordingly, inhibition of the activity of one or more of the enzymes described  
5 herein could lead to inhibition of plant growth. Thus, the instant polypeptides could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant  
10 breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al.  
15 (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map  
20 previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology  
25 outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In:  
30 *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of  
35 large clones (several to several hundred KB; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include

allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptide. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptide can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

### EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean, and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean, and Wheat

Library	Tissue	Clone
cbn10	Corn Developing Kernel (Embryo and Endosperm); 10 Days After Pollination	cbn10.pk0005.e5
cs1	Corn Leaf Sheath From 5 Week Old Plant	cs1.pk0038.g4
p0005	Corn Immature Ear	p0005.cbmev33r
rl0n	Rice 15 Day Old Leaf*	rl0n.pk0062.f12
scr1c	Soybean ( <i>Glycine max</i> L., 2872) Embryogenic Suspension Culture Subjected to 4 Vacuum Cycles and Collected 12 Hrs Later	scr1c.pk004.c14
sdp2c	Soybean Developing Pod (6-7 mm)	sdp2c.pk013.h24
sdp3c	Soybean Developing Pod (8-9 mm)	sdp3c.pk005.h18
sfl1	Soybean Immature Flower	sfl1.pk0072.c4 sfl1.pk0086.d7
sl2	Soybean Two-Week-Old Developing Seedling Treated With 2.5 ppm Chlorimuron	sl2.pk129.112
src3c	Soybean 8 Day Old Root Infected With Cyst Nematode	src3c.pk022.d15
wre1n	Wheat Root From 7 Day Old Etiolated Seedling*	wre1n.pk191.g3

\*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

10

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer

20

sequencing reactions to generate partial cDNA sequences (expressed sequence tags or “ESTs”; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

#### EXAMPLE 2

##### 5                    Identification of cDNA Clones

cDNA clones encoding phospholipid biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) searches for similarity to sequences contained in the BLAST “nr” database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the “nr” database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the “nr” database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as “pLog” values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST “hit” represent homologous proteins.

#### EXAMPLE 3

##### Characterization of cDNA Clones Encoding Phosphatidylglycerophosphate Synthase

25            The BLASTX search using the nucleotide sequences from the contig assembled of clones p0005.cbmev33r and cs1.pk0038.g4 and the contig assembled of clones sdp2c.pk013.h24 and sfl1.pk0086.d7 revealed similarity of the proteins encoded by the cDNAs to phosphatidylglycerophosphate synthase from *Arabidopsis thaliana* and *Pseudomonas fluorescens* (NCBI General Identifier No. 3402692 and 1172467, respectively). The BLAST results for each of these sequences are shown in Table 3:

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous  
to Phosphatidylglycerophosphate Synthase

Clone	BLAST pLog Score	
	3402692	1172467
Contig of: p0005.cbmev33r cs1.pk0038.g4	89.30	35.22
Contig of: sdp2c.pk013.h24 sfl1.pk0086.d7	53.40	15.15

- 5 The sequence of a contig assembled of portion of the cDNA insert from clone p0005.cbmev33r and the entire cDNA insert from clone cs1.pk0038.g4 is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:8. The amino acid sequence set forth in SEQ ID NO:8 was evaluated by BLASTP, yielding a pLog value of 78.52 versus the *Arabidopsis thaliana* sequence and 33.70 versus the *Pseudomonas fluorescens* sequence. The sequence in SEQ ID NO:8 is 71.8% identical to the *Arabidopsis thaliana* sequence and 45.9% identical to the *Pseudomonas fluorescens* sequence. The sequence of a contig assembled of a portion of the cDNA insert from clone sdp2c.pk013.h24 and the entire cDNA insert from clone sfl1.pk0086.d7 is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:10. BLAST scores and
- 10
- 15 probabilities indicate that the instant nucleic acid fragments encode portions of corn and soybean phosphatidylglycerophosphate synthase. These sequences represent the first plant sequences encoding phosphatidylglycerophosphate synthase.

The BLASTX search using the sequences from clones listed in Table 4 revealed similarity of the polypeptides encoded by the cDNAs to phosphatidylglycerophosphate synthase from *Arabidopsis thaliana* (NCBI GenBank Identifier (GI) No. 3402692). Shown in Table 4 is the BLAST result for a contig assembled from an FIS and two ESTs ("Contig\*"):

20

TABLE 4

BLAST Result for Sequence Encoding Polypeptide Homologous  
to Phosphatidylglycerophosphate Synthase

Clone	Status	BLAST pLog Score
		NCBI GenBank Identifier (GI) No. 3402692
Contig of sfl1.pk0086.d7 (FIS) sdp2c.pk013.h24 scr1c.pk004.c14	Contig*	88.52

The data in Table 5 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:8, 10, and 12 and the *Arabidopsis thaliana* sequence (NCBI GenBank Identifier (GI) No. 3402692).

5

TABLE 5

Percent Identity of Amino Acid Sequences  
Deduced From the Nucleotide Sequences of cDNA Clones Encoding  
Polypeptides Homologous to Phosphatidylglycerophosphate Synthase

SEQ ID NO.	Percent Identity to NCBI GenBank Identifier (GI) No. 3402692
8	68.8
10	67.1
12	64.7

10

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a phosphatidylglycerophosphate synthase. These sequences represent the first plant sequences encoding phosphatidylglycerophosphate synthase.

20

EXAMPLE 4Characterization of cDNA Clones Encoding Phosphatidylinositol Synthase

The BLASTX search using the EST sequences from clones cbn10.pk0005.e5 and rl0n.pk0062.f12 and the contig assembled from clones sl2.pk129.112 and src3c.pk022.d15 revealed similarity of the proteins encoded by the cDNAs to phosphatidylinositol synthase from *Arabidopsis thaliana* and *Saccharomyces cerevisiae* (NCBI General Identifier No. 3367632 and 130240, respectively). The BLAST results for each of these ESTs and the contig are shown in Table 6:

25

TABLE 6  
BLAST Results for Clones Encoding Polypeptides Homologous  
to Phosphatidylinositol Synthase

Clone	BLAST pLog Score	
	3367632	130240
cbn10.pk0005.e5	82.30	27.30
rl0n.pk0062.f12	93.10	33.70
Contig of:	36.70	12.40
sl2.pk129.112		
src3c.pk022.d15		

5       The sequence of the entire cDNA insert in clone cbn10.pk0005.e5 was determined and is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:14. The amino acid sequence set forth in SEQ ID NO:14 was evaluated by BLASTP, yielding a pLog value of 79.00 versus the *Arabidopsis thaliana* sequence and a pLog value of 30.22 versus the *Saccharomyces cerevisiae* sequence. The sequence of the

10   entire cDNA insert in clone rl0n.pk0062.f12 was determined and is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:16. The amino acid sequence set forth in SEQ ID NO:16 was evaluated by BLASTP, yielding a pLog value of 79.00 versus the *Arabidopsis thaliana* sequence and a pLog value of 30.22 versus the *Saccharomyces cerevisiae* sequence. The sequence of a contig assembled of a

15   portion of the cDNA insert from clones sl2.pk129.112 and src3c.pk022.d15 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:18. The amino acid sequence presented in SEQ ID NO:14 is 43.3% identical to the *Saccharomyces cerevisiae* sequence and the amino acid sequence presented in SEQ ID NO:16 is 40.3% identical to the *Saccharomyces cerevisiae* sequence.

20       Sequence alignments and percent identity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS*. 5: 151-153) with the default parameters (GAP PENALTY=10, GAP LENGHT PENALTY=10). Sequence

25   percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the default settings (gap penalty=11, gap length penalty=3).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire corn and rice phosphatidylinositol synthase and portion of soybean phosphatidylinositol synthase. These sequences represent the first plant

30   sequences encoding phosphatidylinositol synthase.

The BLASTX search using the EST sequences from the clone listed in Table 7 revealed similarity of the polypeptide encoded by the cDNA to phosphatidylinositol

synthase from *Arabidopsis thaliana* (NCBI GenBank Identifier (GI) No. 4467144). Shown in Table 7 is the BLAST result for a sequence encoding the entire protein derived from an FIS ("CGS"):

5

TABLE 7

BLAST Result for Sequence Encoding Polypeptide Homologous  
to Phosphatidylinositol Synthase

Clone	Status	BLAST pLog Score NCBI GenBank Identifier (GI) No. 4467144
sl2.pk129.112 (FIS)	CGS	102.00

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:14, 16, and 20 and the *Arabidopsis thaliana* sequence (NCBI GenBank Identifier (GI) No. 3367632) (SEQ ID NO:21). The data in Table 8 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:14, 16, and 20 and the *Arabidopsis thaliana* sequence (NCBI GenBank Identifier (GI) No. 3367632) (SEQ ID NO:21).

15

TABLE 8

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences  
of cDNA Clones Encoding Polypeptides Homologous to Phosphatidylinositol Synthase

SEQ ID NO.	Percent Identity to NCBI GenBank Identifier (GI) No. 3367632
14	64.2
16	61.0
20	75.3

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode all or a substantial portion of a phosphatidylinositol synthase. These sequences represent the first plant sequences encoding phosphatidylinositol synthase.

EXAMPLE 5Characterization of cDNA Clones Encoding Other  
CDP-Alcohol Phosphatidyltransferase

The BLASTX search using the EST sequences from clones listed in Table 9 revealed  
 5 similarity of the polypeptides encoded by the cDNAs to a CDP-alcohol phosphatidyl-  
 transferase from *Arabidopsis thaliana* (NCBI GenBank Identifier (GI) No. 4115948).  
 Shown in Table 9 are the BLAST results for individual ESTs ("EST"), the sequences of the  
 entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from  
 two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs  
 10 ("Contig\*"), or sequences encoding the entire protein derived from an FIS, a contig, or an  
 FIS and PCR ("CGS"):

TABLE 9

15 BLAST Results for Sequences Encoding Polypeptides Homologous  
 to CDP-Alcohol Phosphatidyltransferase

Clone	Status	BLAST pLog Score NCBI GenBank Identifier (GI) No. 4115948
sfl1.pk0072.c4	FIS	47.30
Contig of sdp3c.pk005.h18 sfl1.pk0072.c4 (FIS)	CGS	61.51
wre1n.pk191.g3	FIS	37.00

Figure 2 presents an alignment of the amino acid sequence set forth in SEQ ID NO:4  
 and the *Arabidopsis thaliana* sequence NCBI GenBank Identifier (GI) No. 4115948 (SEQ  
 ID NO:22). The data in Table 10 represents a calculation of the percent identity of the  
 20 amino acid sequence set forth in SEQ ID NO:4 and the *Arabidopsis thaliana* sequence NCBI  
 GenBank Identifier (GI) No. 4115948 (SEQ ID NO:22).

TABLE 10

25 Percent Identity of Amino Acid Sequence Deduced From the Nucleotide Sequence  
 of cDNA Clone Encoding Polypeptide Homologous to CDP-Alcohol  
 Phosphatidyltransferase

SEQ ID NO.	Percent Identity to NCBI GenBank Identifier (GI) No. 4115948
4	44.7

Sequence alignments and percent identity calculations were performed using the  
 Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.,  
 30 Madison, WI). Multiple alignment of the sequences was performed using the Clustal  
 method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default

parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones  
5 encode all or a substantial portion of a CDP-alcohol phosphatidyltransferase. These sequences represent the first plant sequences encoding CDP-alcohol phosphatidyltransferase.

#### EXAMPLE 6

##### Expression of Chimeric Genes in Monocot Cells

10 A chimeric gene comprising a cDNA encoding the instant polypeptide in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites  
15 (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the  
20 plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sall-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-Sall fragment from the 3' end of the maize 10 kD zein gene in the vector  
25 pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit;  
30 U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptide, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses  
35 derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable

embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

5       The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

10       The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1  $\mu$ m in diameter) are coated with DNA using the following technique. Ten  $\mu$ g of plasmid DNAs are added to 50  $\mu$ L of a suspension of gold particles (60 mg per mL). Calcium chloride (50  $\mu$ L of a 2.5 M solution) and spermidine free base (20  $\mu$ L of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200  $\mu$ L of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30  $\mu$ L of ethanol. An aliquot (5  $\mu$ L) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

20       For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

30       Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-

supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the  
5 tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

### EXAMPLE 7

#### Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription  
10 terminator from the gene encoding the  $\beta$  subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the  
15 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be  
20 incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising  
25 sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic  
30 embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into  
35 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent

No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptide and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL CaCl<sub>2</sub> (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

#### EXAMPLE 8

##### Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter

system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptide are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

### EXAMPLE 9

#### Evaluating Compounds for Their Ability to Inhibit the Activity of Phospholipid Biosynthetic Enzymes

The polypeptides described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 8, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant polypeptides may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("His<sub>6</sub>"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant polypeptides, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant polypeptides are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant polypeptides may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a (His)<sub>6</sub> peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include  $\beta$ -mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant polypeptides disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, 5 assays for phosphatidylglycerophosphate synthase are presented by S. A. Minskoff and M. L. Greenberg, (1997) *Biochim Biophys Acta* 1348:187-191. Assays for phosphatidylinositol synthase are presented by J. Nikawa and S. Yamashita (1997) *Biochim Biophys Acta* 1348:173-178.

10 Various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

The disclosure of each reference set forth above is incorporated herein by reference in its entirety.

## CLAIMS

What is claimed is:

1. An isolated polynucleotide comprising a first nucleotide sequence encoding a polypeptide of at least 50 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of  
5 SEQ ID NOs:2 and 6; a second nucleotide sequence encoding a polypeptide of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:4; or a third nucleotide sequence comprising the complement of (a) or (b).
- 10 2. The isolated polynucleotide of Claim 1, wherein the first nucleotide sequence consists of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, and 5 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, and 6.
- 15 3. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are DNA.
4. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are RNA.
5. A chimeric gene comprising the isolated polynucleotide of Claim 1 operably linked to suitable regulatory sequences.
- 20 6. An isolated host cell comprising the chimeric gene of Claim 5.
7. A host cell comprising an isolated polynucleotide of Claim 1.
8. The host cell of Claim 7 wherein the host cell is selected from the group consisting of yeast, bacteria, plant, and virus.
9. A virus comprising the isolated polynucleotide of Claim 1.
- 25 10. A polypeptide comprising a first sequence of at least 50 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2 and 6 and a second sequence of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:4.
- 30 11. An isolated polynucleotide comprising a first nucleotide sequence encoding a polypeptide of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:10 and 12; a second nucleotide sequence encoding a polypeptide of at least 150 amino acids that has at least 80% identity based on the Clustal method of alignment  
35 when compared to a polypeptide of SEQ ID NO:8; or a third nucleotide sequence comprising the complement of the first or the second nucleotide sequence.
12. The isolated polynucleotide of Claim 11, wherein the first nucleotide sequence consists of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:7, 9,

and 11 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:8, 10, and 12.

13. The isolated polynucleotide of Claim 11 wherein the nucleotide sequences are DNA.

5 14. The isolated polynucleotide of Claim 11 wherein the nucleotide sequences are RNA.

15. A chimeric gene comprising the isolated polynucleotide of Claim 11 operably linked to suitable regulatory sequences.

16. An isolated host cell comprising the chimeric gene of Claim 15.

10 17. A host cell comprising an isolated polynucleotide of Claim 11.

18. The host cell of Claim 17 wherein the host cell is selected from the group consisting of yeast, bacteria, plant, and virus.

19. A virus comprising the isolated polynucleotide of Claim 11.

20. A polypeptide comprising a first sequence of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:10 and 12 or a second sequence of at least 150 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:8.

21. An isolated polynucleotide comprising a first nucleotide sequence encoding a polypeptide of at least 50 amino acids that has at least 90% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:18; a second nucleotide sequence encoding a polypeptide of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:14; a third nucleotide sequence encoding a polypeptide of at least 150 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:20; a fourth nucleotide sequence encoding a polypeptide of at least 200 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:16; or a fifth nucleotide sequence comprising the complement of (a), (b), (c), or (d).

22. The isolated polynucleotide of Claim 21, wherein the first nucleotide sequence consists of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:13, 15, 17, and 19 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:14, 16, 18, and 20.

23. The isolated polynucleotide of Claim 21 wherein the nucleotide sequences are DNA.

24. The isolated polynucleotide of Claim 21 wherein the nucleotide sequences are RNA.

25. A chimeric gene comprising the isolated polynucleotide of Claim 21 operably linked to suitable regulatory sequences.

26. An isolated host cell comprising the chimeric gene of Claim 25.

27. A host cell comprising an isolated polynucleotide of Claim 21.

5 28. The host cell of Claim 27 wherein the host cell is selected from the group consisting of yeast, bacteria, plant, and virus.

29. A virus comprising the isolated polynucleotide of Claim 21.

30. A polypeptide comprising a first sequence of at least 50 amino acids that has at least 90% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:18; a second sequence of at least 100 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:14; a third sequence of at least 150 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:20; or a fourth sequence of at least 200 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide of SEQ ID NO:16.

31. A method of selecting an isolated polynucleotide that affects the level of expression of a phospholipid biosynthetic enzyme polypeptide in a plant cell, the method comprising the steps of:

20 (a) constructing an isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from an isolated polynucleotide of any of Claims 1, 11, and 21;

(b) introducing the isolated polynucleotide into a plant cell; and

25 (c) measuring the level of a polypeptide in the plant cell containing the polynucleotide to provide a positive selection means.

32. The method of Claim 31 wherein the isolated polynucleotide consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20.

30 33. A method of selecting an isolated polynucleotide that affects the level of expression of a phospholipid biosynthetic enzyme polypeptide in a plant cell, the method comprising the steps of:

(a) constructing an isolated polynucleotide of any of Claims 1, 21, and 31;

(b) introducing the isolated polynucleotide into a plant cell; and

35 (c) measuring the level of polypeptide in the plant cell containing the polynucleotide to provide a positive selection means.

34. A method of obtaining a nucleic acid fragment encoding a phospholipid biosynthetic enzyme polypeptide comprising the steps of:

(a) synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19 and the complement of such sequences; and

5 (b) amplifying a nucleic acid sequence using the oligonucleotide primer.

35. A method of obtaining a nucleic acid fragment encoding a phospholipid biosynthetic enzyme polypeptide comprising the steps of:

10 (a) probing a cDNA or genomic library with an isolated polynucleotide comprising at least one of 30 contiguous nucleotides derived from a sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 and the complement of such sequences;

(b) identifying a DNA clone that hybridizes with the isolated polynucleotide;

(c) isolating the identified DNA clone; and

15 (d) sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

36. A method for evaluating at least one compound for its ability to inhibit the activity of a phospholipid biosynthetic enzyme, the method comprising the steps of:

20 (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a phospholipid biosynthetic enzyme, operably linked to suitable regulatory sequences;

(b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the phospholipid biosynthetic enzyme encoded by the operably linked nucleic acid fragment in the transformed host cell;

25 (c) optionally purifying the phospholipid biosynthetic enzyme expressed by the transformed host cell;

(d) treating the phospholipid biosynthetic enzyme with a compound to be tested; and

30 (e) determining the activity of the phospholipid biosynthetic enzyme that has been treated with the compound, thereby selecting compounds with potential for inhibitory activity.

37. A composition comprising the isolated polynucleotide of any of Claims 1, 11, and 21.

35 38. A composition comprising the isolated polypeptide of any of Claims 10, 20, and 30.

39. An isolated polynucleotide comprising the nucleotide sequence having at least one of 30 contiguous nucleotides derived from a nucleic acid sequence selected from the

group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 and the complement of such sequences.

40. An expression cassette comprising an isolated polynucleotide of any of Claims 1, 11, and 21, operably linked to a promoter.

5 41. A method for positive selection of a transformed cell comprising:

(a) transforming a host cell with the chimeric gene of any of Claims 5, 15, and 25; and

(b) growing the transformed host cell under conditions which allow expression of the polynucleotide in an amount sufficient to complement a yeast *pis* or *pgs1* mutation to provide a positive selection means.

10 42. A method for positive selection of a transformed cell comprising:

(a) transforming a host cell with an expression cassette of Claim 40; and

(b) growing the transformed host cell under conditions which allow expression of the polynucleotide in an amount sufficient to complement a yeast *pis* or *pgs1* mutant to provide a positive selection means.

15 43. The method of Claim 41 wherein the host cell is a plant cell.

44. The method of Claim 43 wherein the plant cell is a monocot.

45. The method of Claim 43 wherein the plant cell is a dicot.

46. The method of Claim 42 wherein the host cell is a plant cell.

20 47. The method of Claim 46 wherein the plant cell is a monocot.

48. The method of Claim 46 wherein the plant cell is a dicot.



SEQ ID NO:04	MVLFKFLKQILVDNTQTKTKARTFFETSTSTATAFPWCAPHNNRFLFRSPGPGP----	* * * * *	* * * * *	* * * * *	60
SEQ ID NO:22	MAIYRSLRKLVEINH---RKTRPFFTAATASGCTVSLTPPQFSPFLPHFHSRLSPLSKW	* * * * *	* * * * *	* * * * *	60
SEQ ID NO:04	-----GPLFLARPPWKLSQSATPLYSHGNA-----FNLLRSTAPLQFPDPVPSPHS	* * * * *	* * * * *	* * * * *	120
SEQ ID NO:22	FVPLNGPLFLSSPPWKLLQSATPLHWRNGSVLKKVEALNRLDRIRSRTFRPRLGLQ	* * * * *	* * * * *	* * * * *	120
SEQ ID NO:04	NKPSIF-----HSEVNLPNFISFTRLSSGPLLAWMISNEFYTSAMVGLA	* * * * *	* * * * *	* * * * *	180
SEQ ID NO:22	VVPNILTVDRNDSKEEDGGKLVKSFVNVPNMISMARLVSGPVLWMMISNEMYSSAFLGLA	* * * * *	* * * * *	* * * * *	180
SEQ ID NO:04	LSGATDW-----LDGYVARKMKIDSVVGSYLDPLADKVLIGCVALAMVHRDLLHPGLVSL	* * * * *	* * * * *	* * * * *	240
SEQ ID NO:22	VSGASDWVCIFFLDGYVARRMKINSVVGSYLDPLADKVLIGCVAVAMVQDKLLHPGLVGI	* * * * *	* * * * *	* * * * *	240
SEQ ID NO:04	VVFRDVLVGGAVFLRANSLGWKWESWDFFNLDGTCRQKVEPLFLSKVNAVQCALVA	* * * * *	* * * * *	* * * * *	300
SEQ ID NO:22	VLLRDVALVGGAVYLRAINLDWK-----VNTVFQLTLVAG	* * * * *	* * * * *	* * * * *	300
SEQ ID NO:04	ALLQPEFGTQETQLYVTYLSYLVASTTVASTAAYGAQYFRRLAVVSNSV	* * * * *	* * * * *	* * * * *	349
SEQ ID NO:22	AILQPEFGNPDTQTWITYLR-----	* * * * *	* * * * *	* * * * *	349

SEQUENCE LISTING

<110> Cahoon, Rebecca E.  
Falco, Carl  
Kinney, Tony

<120> Phospholipid Biosynthetic Enzymes

<130> BB1307

<140>

<141>

<150> 60/112,558

<151> 1998-12-16

<160> 22

<170> Microsoft Office 97

<210> 1

<211> 950

<212> DNA

<213> Glycine max

<400> 1

```
gcacgagctt cgctccacag cccctctcca gttccctgat ccggttcctt cccattccaa 60
caaaccctcg attttccaca gcttcgtcaa tctccccaac ttcatttcct tcacccgatt 120
actttccggg cctctcctcg catggatgat ctccaacgag ttttatactt ctgcaatggg 180
agggctgggt ctctccgggt ccaccgattg gctggatgga tacgtggcta ggaagatgaa 240
gatcgattcc gtagtgggtt cctaccttga tccccttgct gacaagggtc ttattgggtg 300
cgttgctctt gctatgggtc atagagatct actgcatcct ggacttggtt gtcttggttg 360
gttccgggac gtcttcctcg ttggtggcgc agtatttcta agagcaaata gcttgggttg 420
gaagtgggaa agctgggttg atttttttaa ccttgatggg acctgccgcc aaaagggttg 480
accactcttt ctaagcaagg tgaatgcagt gtgccaatta gcactagtgt ctgcagctct 540
tcttcaacca gagtttgaa cccaggaaac tcaattatat gttacttatt tgagctatct 600
agtggcttca acaactgtag catctactgc agcatatgga gcacaatatt ttaggagact 660
ggctgtagta tcaaaactcg tctagttaag tttccattga gatgcagagg gatcagggat 720
cggcatgagc aaggacttat tgcacagtaa attatggcta gattgcacag tctttgtttt 780
caaactttgc cttctaacat aagcattaat gtttcaaata ttttcaagta gtggctgaca 840
aatgttagat tatttttcct gcaattctat gttagaggaa cgtatgaaac gtgtctttct 900
ctttcatgcc atttgaattt atttgatcaa aaaaaaaaaa aaaaaaaaaa 950
```

<210> 2

<211> 227

<212> PRT

<213> Glycine max

<400> 2

```
His Glu Leu Arg Ser Thr Ala Pro Leu Gln Phe Pro Asp Pro Val Pro
  1              5              10              15
```

```
Ser His Ser Asn Lys Pro Ser Ile Phe His Ser Phe Val Asn Leu Pro
      20              25              30
```

```
Asn Phe Ile Ser Phe Thr Arg Leu Ser Gly Pro Leu Leu Ala Trp
      35              40              45
```

```
Met Ile Ser Asn Glu Phe Tyr Thr Ser Ala Met Val Gly Leu Ala Leu
      50              55              60
```

Ser Gly Ala Thr Asp Trp Leu Asp Gly Tyr Val Ala Arg Lys Met Lys  
 65 70 75 80  
 Ile Asp Ser Val Val Gly Ser Tyr Leu Asp Pro Leu Ala Asp Lys Val  
 85 90 95  
 Leu Ile Gly Cys Val Ala Leu Ala Met Val His Arg Asp Leu Leu His  
 100 105 110  
 Pro Gly Leu Val Ser Leu Val Val Phe Arg Asp Val Phe Leu Val Gly  
 115 120 125  
 Gly Ala Val Phe Leu Arg Ala Asn Ser Leu Gly Trp Lys Trp Glu Ser  
 130 135 140  
 Trp Phe Asp Phe Phe Asn Leu Asp Gly Thr Cys Arg Gln Lys Val Glu  
 145 150 155 160  
 Pro Leu Phe Leu Ser Lys Val Asn Ala Val Cys Gln Leu Ala Leu Val  
 165 170 175  
 Ala Ala Ala Leu Leu Gln Pro Glu Phe Gly Thr Gln Glu Thr Gln Leu  
 180 185 190  
 Tyr Val Thr Tyr Leu Ser Tyr Leu Val Ala Ser Thr Thr Val Ala Ser  
 195 200 205  
 Thr Ala Ala Tyr Gly Ala Gln Tyr Phe Arg Arg Leu Ala Val Val Ser  
 210 215 220

Asn Ser Val  
 225

<210> 3  
 <211> 1223  
 <212> DNA  
 <213> Glycine max

<400> 3  
 ccaatctgtg tcgaaaaaatg tttcaataat ggttctgttc cgcaagttcc taaaacaaat 60  
 actcgtcgac aacacccaaa ccaaaacaaa agcccgaacc tttttcacct caacctcaac 120  
 cgcaaccgca ttcccatggt gcgcgcctca ccataacaac cgcctcttcc ggtccccagg 180  
 cccaggccca ggcccaggcc ctctcttccct cgctcggcct ccatggaaac tctctcagtc 240  
 cgccacccct ctctattcgc acggaaacgc cttcaatttg cttcgcctcca cagccccctct 300  
 ccagttccct gatccggttc cctcccattc caacaaaccc tcgattttcc acagcttcgt 360  
 caatctcccc aacttcattt ccttcacccg attactttcc ggtcctctcc tcgcatggat 420  
 gatctccaac gagttttata cttctgcaat ggtagggctg gctctctccg gtgccaccga 480  
 ttggctggat ggatacgtgg ctaggaagat gaagatcgat tccgtagtgg gttcctacct 540  
 tgatccccctt gctgacaagg tccttattgg ttgcgttgct cttgctatgg tgcatagaga 600  
 tctactgcat cctggacttg ttagtcttgt tgtgttccgg gacgtcttcc tcgttggtgg 660  
 cgcagtattt ctaagagcaa atagcttggg ttggaagtgg gaaagctggg ttgatttttt 720  
 taaccttgat gggacctgcc gccaaaaggt tgaaccactc tttctaagca aggtgaatgc 780  
 agtgtgccaa ttagcactag ttgctgcagc tcttcttcaa ccagagtttg gaacccagga 840  
 aactcaatta tatgttactt atttgagcta tttagtggct tcaacaactg tagcatctac 900  
 tgcagcatat ggagcacaat attttaggag actggctgta gtatcaaact cggcttagtt 960  
 aagtttccat tgagatgcag agggatcagg gatcggcatg agcaaggact tattgcacag 1020  
 taaattatgg ctagattgca cagtctttgt tttcaaactt tgccttctaa cataagcatt 1080  
 aatgtttcaa atgttttcaa gtagtggctg acaaagtta gattatTTTT cctgcaattc 1140

tatgttagag gaacgtatga aacgtgtctt tctctttcat gccatttgaa tttatttgat 1200  
 caaaaaaaaa aaaaaaaaaaaa aaa 1223

<210> 4  
 <211> 309  
 <212> PRT  
 <213> Glycine max

<400> 4  
 Met Val Leu Phe Arg Lys Phe Leu Lys Gln Ile Leu Val Asp Asn Thr  
           1                          5                          10                          15  
 Gln Thr Lys Thr Lys Ala Arg Thr Phe Phe Thr Ser Thr Ser Thr Ala  
                           20                          25                          30  
 Thr Ala Phe Pro Trp Cys Ala Pro His His Asn Asn Arg Leu Phe Arg  
                           35                          40                          45  
 Ser Pro Gly Pro Gly Pro Gly Pro Gly Pro Leu Phe Leu Ala Arg Pro  
           50                          55                          60  
 Pro Trp Lys Leu Ser Gln Ser Ala Thr Pro Leu Tyr Ser His Gly Asn  
           65                          70                          75                          80  
 Ala Phe Asn Leu Leu Arg Ser Thr Ala Pro Leu Gln Phe Pro Asp Pro  
                           85                          90                          95  
 Val Pro Ser His Ser Asn Lys Pro Ser Ile Phe His Ser Phe Val Asn  
                           100                          105                          110  
 Leu Pro Asn Phe Ile Ser Phe Thr Arg Leu Leu Ser Gly Pro Leu Leu  
           115                          120                          125  
 Ala Trp Met Ile Ser Asn Glu Phe Tyr Thr Ser Ala Met Val Gly Leu  
           130                          135                          140  
 Ala Leu Ser Gly Ala Thr Asp Trp Leu Asp Gly Tyr Val Ala Arg Lys  
           145                          150                          155                          160  
 Met Lys Ile Asp Ser Val Val Gly Ser Tyr Leu Asp Pro Leu Ala Asp  
                           165                          170                          175  
 Lys Val Leu Ile Gly Cys Val Ala Leu Ala Met Val His Arg Asp Leu  
                           180                          185                          190  
 Leu His Pro Gly Leu Val Ser Leu Val Val Phe Arg Asp Val Phe Leu  
           195                          200                          205  
 Val Gly Gly Ala Val Phe Leu Arg Ala Asn Ser Leu Gly Trp Lys Trp  
           210                          215                          220  
 Glu Ser Trp Phe Asp Phe Phe Asn Leu Asp Gly Thr Cys Arg Gln Lys  
           225                          230                          235                          240  
 Val Glu Pro Leu Phe Leu Ser Lys Val Asn Ala Val Cys Gln Leu Ala  
                           245                          250                          255  
 Leu Val Ala Ala Ala Leu Leu Gln Pro Glu Phe Gly Thr Gln Glu Thr  
                           260                          265                          270

Gln Leu Tyr Val Thr Tyr Leu Ser Tyr Leu Val Ala Ser Thr Thr Val  
275 280 285

Ala Ser Thr Ala Ala Tyr Gly Ala Gln Tyr Phe Arg Arg Leu Ala Val  
290 295 300

Val Ser Asn Ser Val  
305

<210> 5  
<211> 705  
<212> DNA  
<213> Triticum aestivum

<400> 5  
gcacgagcga tcgggcgcat ggcgtcgggt ccgctcattg gatggatgat tatgaatgaa 60  
tggatatcttc ctgcttttgg cacattagct ttgtctggtg caagtgattg gcaagaaaga 120  
tgggcatcaa ttctgtaatt ggttcatatc tggacccatt ggctgacaag gtgttgattt 180  
gctgtgttgc tgtagcaatg gttgagaggg agctcttaca tcctggcctt gttggcctgg 240  
tggttgcaag agacttttct ctggtgggtg gggcttttcta caaacgagca tccagcctgg 300  
gatggaagtg gaatagtttg tcagactttt gtaatttggg tgcaattcag cgtgagaagg 360  
ttcaacctct tttcatcagc aaggtaaata cagtatttca gttgatgttg gttgctggcg 420  
cactccttca gccagaattc ggcacagagg agatgcgga ttatattaca tacttggtt 480  
ggctggttgc tactactaca atggcatcca cagtaggtta tgggtgtgaaa tactatcagt 540  
taagaccgag gacatgatga cagctgttgc ctgcattttt gtttaattacc acccagatta 600  
cgattggtga atccagcaat tagttgtgaa actcaattgt ttccatggcc catccaaatt 660  
tgaagaaatg gaagggttgt gcagtttcta aaaaaaaaaa aaaaa 705

<210> 6  
<211> 149  
<212> PRT  
<213> Triticum aestivum

<400> 6  
Leu Ala Arg Lys Met Gly Ile Asn Ser Val Ile Gly Ser Tyr Leu Asp  
1 5 10 15  
Pro Leu Ala Asp Lys Val Leu Ile Cys Cys Val Ala Val Ala Met Val  
20 25 30  
Glu Arg Glu Leu Leu His Pro Gly Leu Val Gly Leu Val Val Ala Arg  
35 40 45  
Asp Phe Leu Leu Val Gly Gly Ala Phe Tyr Lys Arg Ala Ser Ser Leu  
50 55 60  
Gly Trp Lys Trp Asn Ser Trp Ser Asp Phe Cys Asn Leu Asp Ala Ile  
65 70 75 80  
Gln Arg Glu Lys Val Gln Pro Leu Phe Ile Ser Lys Val Asn Thr Val  
85 90 95  
Phe Gln Leu Met Leu Val Ala Gly Ala Leu Leu Gln Pro Glu Phe Gly  
100 105 110  
Thr Glu Glu Met Arg Asn Tyr Ile Thr Tyr Leu Gly Trp Leu Val Ala  
115 120 125  
Thr Thr Thr Met Ala Ser Thr Val Gly Tyr Gly Val Lys Tyr Tyr Gln  
130 135 140

Leu Arg Pro Arg Thr  
145

<210> 7  
<211> 1109  
<212> DNA  
<213> Zea mays

<220>  
<221> unsure  
<222> (20)

<400> 7  
ggcacgaggg gcaggaagan gcggtgcca gcggtgaacg ggctcgccaa ggacgcgcct 60  
gcggtgaacg ggatgcgcaa ggaggagccc gtggccgcgc ctcccagggt gctcaggctg 120  
cccactgtgc tcaccattgg gcgcgtcgct gctgtgcctc ttctgatcag cactttctac 180  
atggagggtc catgggcagc tatgggcaca acaggcatct tccttgctgc cgcagtcact 240  
gattggctag atggttatat tgctagaaag atgcagctag gaacaccttt tgggtgcat 300  
cttgatcctg tggccgacaa gcttatggta gcggaacat tagttttgct atgcaccaa 360  
cctttgggat catcactgct caatgatggg ccatgggttc taacagtccc ttccattgcc 420  
atcattggga gagagattac aatgtcagct gtgagagagt gggctgcac tcagaataac 480  
aaagttcttg aggtgtgtgc agttaacaat ttaaggaagt ggaagacagc gacgcagatg 540  
acagcattga ctctgcttct cgcaagcaga gacccaagtc tacctgcgca aggcgtcta 600  
gttgcccccg gtgttgctct actttacgtc cccacgggac tcgccgtatg gtccctagt 660  
gtgtacatga gaaagatatg gcggatactg ctcaaatagt attagcaca tatatatata 720  
gtgtaataga agaaacaagc aaggcgagca tccgttcgtg atcgtacaac ttttggagg 780  
ctgttggtg gcctgagccc tagagctggt tgggatcctc attcatctgt aactgtg 840  
tcagtttttg gtggatttgc cgctaaattt gctgaggcat gccaggatcc tattagatt 900  
ttagaaacaa gacatggtgt gaccctccgt gcgtaatgag aattcttaga aacaagaaga 960  
catggtgtgg ccctgcgtgc gtaatgagat actagtctgc gatgcagaag attgtaacct 1020  
tatacaaata caatctggag gtttctgcgt gtggagttct accgttgtac tctacacata 1080  
atatcataat aaaaaaaaaa aaaaaaaaaa 1109

<210> 8  
<211> 208  
<212> PRT  
<213> Zea mays

<400> 8  
Met Arg Lys Glu Glu Pro Val Ala Ala Pro Pro Arg Leu Leu Arg Leu  
1 5 10 15  
Pro Thr Val Leu Thr Ile Gly Arg Val Ala Ala Val Pro Leu Leu Ile  
20 25 30  
Ser Thr Phe Tyr Met Glu Gly Pro Trp Ala Ala Met Gly Thr Thr Gly  
35 40 45  
Ile Phe Leu Ala Ala Ala Val Thr Asp Trp Leu Asp Gly Tyr Ile Ala  
50 55 60  
Arg Lys Met Gln Leu Gly Thr Pro Phe Gly Ala Phe Leu Asp Pro Val  
65 70 75 80  
Ala Asp Lys Leu Met Val Ala Ala Thr Leu Val Leu Leu Cys Thr Lys  
85 90 95  
Pro Leu Gly Ser Ser Leu Leu Asn Asp Gly Pro Trp Val Leu Thr Val  
100 105 110

Pro Ser Ile Ala Ile Ile Gly Arg Glu Ile Thr Met Ser Ala Val Arg  
115 120 125

Glu Trp Ala Ala Ser Gln Asn Asn Lys Val Leu Glu Ala Val Ala Val  
130 135 140

Asn Asn Leu Arg Lys Trp Lys Thr Ala Thr Gln Met Thr Ala Leu Thr  
145 150 155 160

Leu Leu Leu Ala Ser Arg Asp Pro Ser Leu Pro Ala Gln Gly Ala Leu  
165 170 175

Val Ala Pro Gly Val Ala Leu Leu Tyr Val Pro Thr Gly Leu Ala Val  
180 185 190

Trp Ser Leu Val Val Tyr Met Arg Lys Ile Trp Arg Ile Leu Leu Lys  
195 200 205

<210> 9

<211> 826

<212> DNA

<213> Glycine max

<400> 9

```

atgaaatcta catttggtgc ctttttagat ccagtagctg acaagcttat ggtttctgcc 60
acattggtct tattatgtac tagaccgttg gaagttgctg tgtttcgaca agtaccatgg 120
ctattgatta taccttcaat caccataatt ggcagagaga taaccatgtc ggcactcagg 180
gaatgggctg cttcccagg tagcaagctt ctggagggtg ttgcagttaa taatttgggg 240
aaatggaaaa cagcatcaca gatgggtggcg ttaatcatcc tccttgctac ccgggactgc 300
agtcgtggag gacctgctat tttggtggga tctggggttg tcttgcttta catttccgca 360
gggcttgccct tatggtcatt tgtagtatat atgagagaaa tatgcaagggt gttgcggagg 420
tagcatgttt gattgctctc tgccttcaca tatatccgct tgtttggtgg gttttattct 480
tataatgttt gctctccaca atgtctaact ctgcaaagca ccattcctcc tattcttgaa 540
aaatgtattg tttttccccc ttagttacca ctggtaaagg aaagatggaa aaaaagaaag 600
aaagatgaca tgtggattga attggaagaa ttcagggttac tagtcacgcg aatcgaattc 660
ctatctacag tttggagcct attaaagttt gctacctgag aaatttcgcc agaggaatca 720
atgtatgggt ctagttgact cttaaattga tttttgtcct ggtcatatta tttgccaata 780
cttaaataca acagttattt ttcaaaaaaa aaaaaaaaaa aaaaaa 826

```

<210> 10

<211> 140

<212> PRT

<213> Glycine max

<400> 10

Met Lys Ser Thr Phe Gly Ala Phe Leu Asp Pro Val Ala Asp Lys Leu  
1 5 10 15

Met Val Ser Ala Thr Leu Val Leu Leu Cys Thr Arg Pro Leu Glu Val  
20 25 30

Ala Val Phe Arg Gln Val Pro Trp Leu Leu Ile Ile Pro Ser Ile Thr  
35 40 45

Ile Ile Gly Arg Glu Ile Thr Met Ser Ala Leu Arg Glu Trp Ala Ala  
50 55 60

Ser Gln Gly Ser Lys Leu Leu Glu Val Val Ala Val Asn Asn Leu Gly  
65 70 75 80

Lys Trp Lys Thr Ala Ser Gln Met Val Ala Leu Ile Ile Leu Leu Ala  
85 90 95

Thr Arg Asp Cys Ser Arg Gly Gly Pro Ala Ile Leu Val Gly Ser Gly  
100 105 110

Val Val Leu Leu Tyr Ile Ser Ala Gly Leu Ala Leu Trp Ser Phe Val  
115 120 125

Val Tyr Met Arg Glu Ile Cys Lys Val Leu Arg Arg  
130 135 140

<210> 11  
<211> 1149  
<212> DNA  
<213> Glycine max

<400> 11  
catatggttg ccgagatgaa tgtccgtgaa ttttccatct ctaagaatga ccgtacttca 60  
aattccaatt ccaaatttct taccttgccc accattttga ctcttggccg tgtcgccgct 120  
gtacctcttc ttgttgccac cttctacatg gatggttggc gaggaacagt tgttactaca 180  
agtatcttca ctgctgcagc agttacagat tggcttgatg gttatattgc tcgtaagatg 240  
aaaatgaaat ctacatttgg tgccttttta gatccagtag ctgacaagct tatggtttct 300  
gccacattgg tcttattatg tactagaccg ttggaagtig ctgtgtttcg acaagtacca 360  
tggctattga ttataccttc aatcaccata attggcagag agataacat gtcggcactc 420  
agggaatggg ctgcttccca gggtagcaag cttctggagg ttgttgagc taataatttg 480  
gggaaatgga aaacagcatc acagatgggtg gcgttaatca tcctccttgc taccggggac 540  
tgcagtcgtg gaggacctgc tattttgggtg ggatctgggg ttgtcttgct ttacatttcc 600  
gcagggccttg ccttatggtc attttagta tatatgagag aaatatgcaa ggtgttgcg 660  
aggtagcatg tttgattgct ctctgccttc acatatatcc gcttggttgg tgggttttat 720  
tcttttgatc atttcttacg taaacagttt agttagtgt tccatggaag ggaatgtagt 780  
ccttgtagt tgtgaccaat atatataatg tttgctctcc acaatgtcta actctgcaaa 840  
gcaccattcc tcctattctt gaaaattgta ttgtttttcc cccttagtta ccaactggtaa 900  
aggaaagatg gaaaaaaaga aagaaagatg acatgtggat tgaattggaa gaattcaggt 960  
tactagtcac gcgaatcgaa ttcttatcta cagtttggag cctattaaag tttgctacct 1020  
gagaaatttc gccagaggaa tcaatgtatg gttctagtig actcttaaat tgatttttgt 1080  
cctggtcata ttatttgcca atacttaaat acaacagtta tttttcaaaa aaaaaaaaaa 1140  
aaaaaaaaa 1149

<210> 12  
<211> 221  
<212> PRT  
<213> Glycine max

<400> 12  
His Met Val Ala Glu Met Asn Val Arg Glu Phe Ser Ile Ser Lys Asn  
1 5 10 15  
Asp Arg Thr Ser Asn Ser Asn Ser Lys Phe Leu Thr Leu Pro Thr Ile  
20 25 30  
Leu Thr Leu Gly Arg Val Ala Ala Val Pro Leu Leu Val Ala Thr Phe  
35 40 45  
Tyr Met Asp Gly Trp Arg Gly Thr Val Val Thr Thr Ser Ile Phe Thr  
50 55 60  
Ala Ala Ala Val Thr Asp Trp Leu Asp Gly Tyr Ile Ala Arg Lys Met  
65 70 75 80

Lys Met Lys Ser Thr Phe Gly Ala Phe Leu Asp Pro Val Ala Asp Lys  
                                     85                                    90                                    95  
 Leu Met Val Ser Ala Thr Leu Val Leu Leu Cys Thr Arg Pro Leu Glu  
                                     100                                    105                                    110  
 Val Ala Val Phe Arg Gln Val Pro Trp Leu Leu Ile Ile Pro Ser Ile  
                                     115                                    120                                    125  
 Thr Ile Ile Gly Arg Glu Ile Thr Met Ser Ala Leu Arg Glu Trp Ala  
                                     130                                    135                                    140  
 Ala Ser Gln Gly Ser Lys Leu Leu Glu Val Val Ala Val Asn Asn Leu  
                                     145                                    150                                    155                                    160  
 Gly Lys Trp Lys Thr Ala Ser Gln Met Val Ala Leu Ile Ile Leu Leu  
                                     165                                    170                                    175  
 Ala Thr Arg Asp Cys Ser Arg Gly Gly Pro Ala Ile Leu Val Gly Ser  
                                     180                                    185                                    190  
 Gly Val Val Leu Leu Tyr Ile Ser Ala Gly Leu Ala Leu Trp Ser Phe  
                                     195                                    200                                    205  
 Val Val Tyr Met Arg Glu Ile Cys Lys Val Leu Arg Arg  
                                     210                                    215                                    220

<210> 13  
 <211> 1258  
 <212> DNA  
 <213> Zea mays

<400> 13  
 gcacgagggc caggtgcctg cattactggg tccgggacct gcgtcagcgc aaccgaacac 60  
 gccgcgacc gctgccgcc ttctcccagc tgcccaggtc ttcgtcgggc cctttgcctg 120  
 cggcgtcggc gacgagcgcc tgccacacca ggaatccgt taagtttatt tatgctcacc 180  
 aatttaaggg cagaacttac ggacctcaag ctcttggtta ctctgatct tctctaggcg 240  
 actaggcaca attgccctct ggcaaaatgc catcagttta tctttacatc cctaacatta 300  
 tcgggtatatt taggatcatc ataaatttca ttgcatttgc ggtttgctat tccaacaagg 360  
 ctctctttgc tatcctgtac ttcatcagct ttgtccttga tgggtgtggat gggttggttg 420  
 caaggaagtt caatcaagca tcaacctttg gagctgtgtt agacatgggt acagataggg 480  
 ttagcactgc ttgtttgttg gcccttctct ccagtttta cagacctggt ttagtcttct 540  
 tgatattgct tggattggat attacgagcc actggtttca aatgtacagt tctttcttgt 600  
 caggtaagac tagccacaag gatgtaaaac acacaggcaa ttggcttctg aaattatatt 660  
 atgggtacag gccattcatg gccttctgct gtgtttcttg tgaggtttta tatattttcc 720  
 tgtttctctt tgctgatgag gagtcaacaa gcttgcttat tgtatgcaa ggcatcctga 780  
 accaaagtcc cgtcgttacc ttggtgtttg ttccactct agttggctgg gcagtgaagc 840  
 aagccaccaa cgtcatccag atgaaaactg ctgcggacgc atgcgtggtg tatgatctga 900  
 agcgagcaa atgaagcatg aaggcagct cagggtttag tatcgacata tccaagggaa 960  
 aactctgcga gggggcgggc tacgtcttgc gtgccttgac atctttctga tgatgcggtc 1020  
 atatgtggga ccaggggatg acatgccgtg gccaatgcaa acaattgttt tgtgaaagca 1080  
 gcggcccgtt aagttgttgt cagtgtgaga gtggtgatgc gatcatgatc ctttttacct 1140  
 agagtagctc ccctttgtgt tagcctgaac gatgttttgc aagccgcatg ttccgaactc 1200  
 taggattatt tggattacaa aacttacata ttccatcctc aaaaaaaaaa aaaaaaaaaa 1258

<210> 14  
 <211> 215  
 <212> PRT  
 <213> Zea mays

&lt;400&gt; 14

Met Pro Ser Val Tyr Leu Tyr Ile Pro Asn Ile Ile Gly Tyr Phe Arg  
 1 5 10 15

Ile Ile Ile Asn Phe Ile Ala Phe Ala Val Cys Tyr Ser Asn Lys Ala  
 20 25 30

Leu Phe Ala Ile Leu Tyr Phe Ile Ser Phe Val Leu Asp Gly Val Asp  
 35 40 45

Gly Trp Phe Ala Arg Lys Phe Asn Gln Ala Ser Thr Phe Gly Ala Val  
 50 55 60

Leu Asp Met Val Thr Asp Arg Val Ser Thr Ala Cys Leu Leu Ala Leu  
 65 70 75 80

Leu Ser Gln Phe Tyr Arg Pro Gly Leu Val Phe Leu Ile Leu Leu Gly  
 85 90 95

Leu Asp Ile Thr Ser His Trp Phe Gln Met Tyr Ser Ser Phe Leu Ser  
 100 105 110

Gly Lys Thr Ser His Lys Asp Val Lys His Thr Gly Asn Trp Leu Leu  
 115 120 125

Lys Leu Tyr Tyr Gly Tyr Arg Pro Phe Met Ala Phe Cys Cys Val Ser  
 130 135 140

Cys Glu Val Leu Tyr Ile Phe Leu Phe Leu Phe Ala Asp Glu Glu Ser  
 145 150 155 160

Thr Ser Leu Leu Ile Val Cys Lys Gly Ile Leu Asn Gln Ser Pro Val  
 165 170 175

Val Ile Leu Val Phe Val Ser Thr Leu Val Gly Trp Ala Val Lys Gln  
 180 185 190

Ala Thr Asn Val Ile Gln Met Lys Thr Ala Ala Asp Ala Cys Val Val  
 195 200 205

Tyr Asp Leu Lys Arg Ser Lys  
 210 215

&lt;210&gt; 15

&lt;211&gt; 1234

&lt;212&gt; DNA

&lt;213&gt; Oryza sativa

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (471)

&lt;400&gt; 15

gcacgagctt acatgcacaa ggaaggaagc atcgaatcga acgacaaatc gaatcaatcc 60  
 atcggcggcg gcggcgactc ggctctcgct tccacgcata catccccgcg cccccctcct 120  
 ctccggcgac tccccggaat cgcggcgcg cggcgcggcc gccctagcgt cggcgacggc 180  
 taggtcttcc tccccgaccc cttcgtgagg tgaggaattc ctatatagga ctacaattta 240  
 agggagaagc ttccaggcct ctagcaagta ataactactg atctccaatg gcacaacctt 300  
 cttctaagaa gacgccgtca gtctatcttt acatccctaa tatcattgga tatttttagga 360

```

tcatcataaa cttcattgct tttgccgtat gttattccaa taggggtgctc tttgctatcc 420
tatacttctt cagctttttc tgcgatggct tggatgggtg gtttgcacga nagtttaacc 480
aagcatcaac atgtggagct gtgctggaca tggtaacaga taggggttagc actgcctggt 540
tggtggcact tctctcccag ttttacagac ctggcttagt tttcttgatg ttgcttgggt 600
tggatattac aagccattgg tttcaaagt acagttcgtt cctatcaggt aagactaccc 660
acaaggatgt aaaagacaca ggcaattggc ttctgaagtt atattatgga caccgaccat 720
tcatggcctt ctgttggtgt gcttcggagg ttctgtacat agttcttttt ctgttcgccg 780
atgagaagtc aacaagcttg cttaatgtgt gcggaacct tctgaagcaa agtcctctca 840
ctgtctttgt tttcatttca actctagttg gttgggcgtt gaaacaagt atcaatgtta 900
tccagatgaa atcagccgca gacgcgtgct ttgtgtttga tttgaagcgc ggcaagtga 960
cagaaatgca caacgatttc gtttagcttg tggttggcat gtcgaagaaa atattccatg 1020
aggcttgatg taccaaggct gtacaaacct aatgtaagat tttgacatgt ccttgacact 1080
gagggagggc atgctgtgca ttaacagcat gaaaagaaaa ggagagagat cagatgtatt 1140
tgaggaaaac agggtttttag tcagattttg tatgtttctt tttacaacag cgatgaatca 1200
gcgcgatctt tgtttcttta tgaaaaaaaa aaaa 1234

```

&lt;210&gt; 16

&lt;211&gt; 223

&lt;212&gt; PRT

&lt;213&gt; Oryza sativa

&lt;220&gt;

&lt;221&gt; UNSURE

&lt;222&gt; (62)

&lt;400&gt; 16

```

Met Ala Gln Pro Ser Ser Lys Lys Thr Pro Ser Val Tyr Leu Tyr Ile
  1                      5                      10                      15

Pro Asn Ile Ile Gly Tyr Phe Arg Ile Ile Ile Asn Phe Ile Ala Phe
                20                      25                      30

Ala Val Cys Tyr Ser Asn Arg Val Leu Phe Ala Ile Leu Tyr Phe Phe
      35                      40                      45

Ser Phe Phe Cys Asp Gly Leu Asp Gly Trp Phe Ala Arg Xaa Phe Asn
    50                      55                      60

Gln Ala Ser Thr Cys Gly Ala Val Leu Asp Met Val Thr Asp Arg Val
    65                      70                      75                      80

Ser Thr Ala Cys Leu Leu Ala Leu Leu Ser Gln Phe Tyr Arg Pro Gly
                85                      90                      95

Leu Val Phe Leu Met Leu Leu Gly Leu Asp Ile Thr Ser His Trp Phe
      100                      105                      110

Gln Met Tyr Ser Ser Phe Leu Ser Gly Lys Thr Thr His Lys Asp Val
    115                      120                      125

Lys Asp Thr Gly Asn Trp Leu Leu Lys Leu Tyr Tyr Gly His Arg Pro
    130                      135                      140

Phe Met Ala Phe Cys Cys Val Ala Ser Glu Val Leu Tyr Ile Val Leu
    145                      150                      155                      160

Phe Leu Phe Ala Asp Glu Lys Ser Thr Ser Leu Leu Asn Val Cys Gly
                165                      170                      175

```

Asn Leu Leu Lys Gln Ser Pro Leu Thr Val Phe Val Phe Ile Ser Thr  
 180 185 190

Leu Val Gly Trp Ala Leu Lys Gln Val Ile Asn Val Ile Gln Met Lys  
 195 200 205

Ser Ala Ala Asp Ala Cys Val Val Phe Asp Leu Lys Arg Gly Lys  
 210 215 220

<210> 17  
 <211> 513  
 <212> DNA  
 <213> Glycine max

<400> 17  
 gttgttgtct ttgaacgtat actacgctaa cttctctttc ccgaatcttt tccaatctcc 60  
 cttcgccgat ccttcgcgcg ttccgccacc acctccgccc ttctcagctc cccgcttccg 120  
 gaaaattatg tatagtgttt tcttggtctg ttcatagtcc tccctccctt gccaccggtg 180  
 aatatttgat tggcaatcct gtccttatta gtcaagtttt agaatacaaa gaagcttcaa 240  
 tggccaagaa accaggacca agatccagta aattgtcagt gtacctttac attcctaata 300  
 taattggata catccgggtt cttttaaact gttttgcctt ctcccaatgt ttatcaaaca 360  
 aaatcttatt ctctattctg tattttttta gctttgtatg tgatgctgtt gatggctggt 420  
 gtgcccgc aa atttaataca gtgtcaacct ttggagctgt gctggacatg gtaacaagac 480  
 aaaattagca ctgcttgccc acttgtaagt ccc 513

<210> 18  
 <211> 79  
 <212> PRT  
 <213> Glycine max

<400> 18  
 Met Ala Lys Lys Pro Gly Pro Arg Ser Ser Lys Leu Ser Val Tyr Leu  
 1 5 10 15

Tyr Ile Pro Asn Ile Ile Gly Tyr Ile Arg Val Leu Leu Asn Cys Phe  
 20 25 30

Ala Phe Ser Gln Cys Leu Ser Asn Lys Ile Leu Phe Ser Ile Leu Tyr  
 35 40 45

Phe Leu Ser Phe Val Cys Asp Ala Val Asp Gly Trp Cys Ala Arg Lys  
 50 55 60

Phe Asn Gln Val Ser Thr Phe Gly Ala Val Leu Asp Met Val Thr  
 65 70 75

<210> 19  
 <211> 1246  
 <212> DNA  
 <213> Glycine max

<400> 19  
 ttttgttgtt gtctttgaac gtataactacg ctaactttctc tttcccgaat cttttccaat 60  
 ctcccttcgc cgatccttcg cgcggtccgc caccacctcc gccgtttctca gctccccgct 120  
 tccggaaaat tatgtatagt gttttcttgg ctgcttcata gtcctccctc ccttgccacc 180  
 ggtgaatatt tgattggcaa tcctgtcctt attagtcaag ttttagaata caaagaagct 240  
 tcaatggcca agaaaccagg accaagatcc agtaaattgt cagtgtacct ttacattcct 300  
 aatataattg gatacatccg ggttctttta aactgttttg ccttctccca atgtttatca 360  
 aacaaaatct tattctctat tctgtatttt ttaagctttg tatgtgatgc tggtgatggc 420  
 tgggtgtgcc gcaaatttaa tcaagtgtca acctttggag ctgtgctgga catggttaaca 480

```

gacagaatta gcaactgcttg tctacttgta gtcctttccc aactgtacaa gcctggcctt 540
agcttcttgt cattgctcgc tttagatatt gccagccact ggctgcaa at gtacagcact 600
ttcttgacgg gaaagactag tcataaagat gtaaaagaca gcagcagctg gcttttcagg 660
gcatactatg gaaataggat gtttatggct tactgctgtg tctcatgtga ggttctttac 720
ttaatcctgt tttatcttgc cgagaatcaa acagagaaac tggtagacgt tatatcaagt 780
aatttacaaa agatatcatt cctttctctt ctaatgggta ctagtattatt tggatgggca 840
gtcaagcaaa ttataaatgt tatccagatg aagacagcag cagatgcgtg tgtgctttat 900
gacatcgaaa aagaacacaa gaactgatgt ttcttccatg accaggatcc ccttgtcttt 960
gtttttcaat accatgtgcg ctcaaagatt aatatgaatt atcccagatg gcaattgggc 1020
ttgaattgtt cagttttcta taatttgtaa tgaaagaaaa ataacatact gaactgaaga 1080
atctaccttt ctatatagca aaaagattct gtagcgcaag gttcagttct tttttgttta 1140
gccattgttt atcacactgc gacgctaaga aagatttggt cagttttgaa attgcaaatc 1200
cttttatagg attggctgat cgtgtactta aaaaaaaaaa aaaaaa 1246

```

&lt;210&gt; 20

&lt;211&gt; 227

&lt;212&gt; PRT

&lt;213&gt; Glycine max

&lt;400&gt; 20

```

Met Ala Lys Lys Pro Gly Pro Arg Ser Ser Lys Leu Ser Val Tyr Leu
  1              5              10              15

Tyr Ile Pro Asn Ile Ile Gly Tyr Ile Arg Val Leu Leu Asn Cys Phe
              20              25              30

Ala Phe Ser Gln Cys Leu Ser Asn Lys Ile Leu Phe Ser Ile Leu Tyr
  35              40              45

Phe Leu Ser Phe Val Cys Asp Ala Val Asp Gly Trp Cys Ala Arg Lys
  50              55              60

Phe Asn Gln Val Ser Thr Phe Gly Ala Val Leu Asp Met Val Thr Asp
  65              70              75              80

Arg Ile Ser Thr Ala Cys Leu Leu Val Val Leu Ser Gln Leu Tyr Lys
              85              90              95

Pro Gly Leu Ser Phe Leu Ser Leu Leu Ala Leu Asp Ile Ala Ser His
              100              105              110

Trp Leu Gln Met Tyr Ser Thr Phe Leu Thr Gly Lys Thr Ser His Lys
              115              120              125

Asp Val Lys Asp Ser Ser Ser Trp Leu Phe Arg Ala Tyr Tyr Gly Asn
              130              135              140

Arg Met Phe Met Ala Tyr Cys Cys Val Ser Cys Glu Val Leu Tyr Leu
              145              150              155              160

Ile Leu Phe Tyr Leu Ala Glu Asn Gln Thr Glu Lys Leu Val Asp Val
              165              170              175

Ile Ser Ser Asn Leu Gln Lys Ile Ser Phe Leu Ser Leu Leu Met Gly
              180              185              190

Thr Ser Leu Phe Gly Trp Ala Val Lys Gln Ile Ile Asn Val Ile Gln
              195              200              205

```

Met Lys Thr Ala Ala Asp Ala Cys Val Leu Tyr Asp Ile Glu Lys Glu  
 210 215 220

His Lys Asn  
 225

<210> 21  
 <211> 227  
 <212> PRT  
 <213> Arabidopsis thaliana

<400> 21

Met Ala Lys Lys Glu Arg Pro Arg Pro Glu Lys Leu Ser Val Tyr Leu  
 1 5 10 15

Tyr Ile Pro Asn Ile Val Gly Tyr Met Arg Val Leu Leu Asn Cys Val  
 20 25 30

Ala Phe Ala Val Cys Phe Ser Asn Lys Pro Leu Phe Ser Val Leu Tyr  
 35 40 45

Phe Phe Ser Phe Cys Cys Asp Ala Val Asp Gly Trp Val Ala Arg Arg  
 50 55 60

Phe Asn Gln Val Ser Thr Phe Gly Ala Val Leu Asp Met Val Thr Asp  
 65 70 75 80

Arg Val Ser Thr Ala Cys Leu Leu Val Ile Leu Ser Gln Ile Tyr Arg  
 85 90 95

Pro Ser Leu Val Phe Leu Ser Leu Leu Ala Leu Asp Ile Ala Ser His  
 100 105 110

Trp Leu Gln Met Tyr Ser Thr Phe Leu Ala Gly Lys Ser Ser His Lys  
 115 120 125

Asp Val Lys Asp Ser Thr Ser Trp Leu Phe Arg Leu Tyr Tyr Gly Asn  
 130 135 140

Arg Ile Phe Met Cys Tyr Cys Cys Val Ser Cys Glu Val Leu Tyr Ile  
 145 150 155 160

Ile Leu Leu Leu Ile Ala Lys Asn Gln Ser Glu Asn Leu Leu Asn Val  
 165 170 175

Val Val Ala Thr Leu Thr Gln Ile Ser Pro Leu Ser Phe Leu Leu Ala  
 180 185 190

Leu Thr Leu Phe Gly Trp Ser Met Lys Gln Thr Ile Asn Val Ile Gln  
 195 200 205

Met Lys Thr Ala Ala Asp Val Cys Val Leu Tyr Asp Ile Glu Lys Gln  
 210 215 220

Gln Lys Pro  
 225

<210> 22  
 <211> 291

&lt;212&gt; PRT

&lt;213&gt; Arabidopsis thaliana

&lt;400&gt; 22

```

Met Ala Ile Tyr Arg Ser Leu Arg Lys Leu Val Glu Ile Asn His Arg
 1           5           10           15

Lys Thr Arg Pro Phe Phe Thr Ala Ala Thr Ala Ser Gly Gly Thr Val
          20           25           30

Ser Leu Thr Pro Pro Gln Phe Ser Pro Leu Phe Pro His Phe Ser His
      35           40           45

Arg Leu Ser Pro Leu Ser Lys Trp Phe Val Pro Leu Asn Gly Pro Leu
      50           55           60

Phe Leu Ser Ser Pro Pro Trp Lys Leu Leu Gln Ser Ala Thr Pro Leu
      65           70           75           80

His Trp Arg Gly Asn Gly Ser Val Leu Lys Lys Val Glu Ala Leu Asn
          85           90           95

Leu Arg Leu Asp Arg Ile Arg Ser Arg Thr Arg Phe Pro Arg Gln Leu
          100          105          110

Gly Leu Gln Ser Val Val Pro Asn Ile Leu Thr Val Asp Arg Asn Asp
      115           120           125

Ser Lys Glu Glu Asp Gly Gly Lys Leu Val Lys Ser Phe Val Asn Val
      130           135           140

Pro Asn Met Ile Ser Met Ala Arg Leu Val Ser Gly Pro Val Leu Trp
      145           150           155           160

Trp Met Ile Ser Asn Glu Met Tyr Ser Ser Ala Phe Leu Gly Leu Ala
          165          170          175

Val Ser Gly Ala Ser Asp Trp Val Cys Ile Phe Phe Leu Asp Gly Tyr
          180          185          190

Val Ala Arg Arg Met Lys Ile Asn Ser Val Val Gly Ser Tyr Leu Asp
          195          200          205

Pro Leu Ala Asp Lys Val Leu Ile Gly Cys Val Ala Val Ala Met Val
      210           215           220

Gln Lys Asp Leu Leu His Pro Gly Leu Val Gly Ile Val Leu Leu Arg
      225           230           235           240

Asp Val Ala Leu Val Gly Gly Ala Val Tyr Leu Arg Ala Leu Asn Leu
          245          250          255

Asp Trp Lys Val Asn Thr Val Phe Gln Leu Thr Leu Val Ala Gly Ala
          260          265          270

Ile Leu Gln Pro Glu Phe Gly Asn Pro Asp Thr Gln Thr Trp Ile Thr
      275           280           285

Tyr Leu Arg
      290

```

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 7 C12N15/54 C12N9/12 C07K14/415 C12N5/10 G01N33/50  
 C12Q1/68 C12Q1/48

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SASAKI ET AL.: "Oryza sativa cDNA, partial sequence (E31409_6A)" EMBL SEQUENCE DATABASE, 16 October 1998 (1998-10-16), XP002136849 HEIDELBERG DE AC AU029728 the whole document	21-24, 30
A	COVITZ ET AL.: "00937 MtrHE Medicago truncatula cDNA 5' similar to phosphatidylinositol synthase, mRNA sequence" EMBL SEQUENCE DATABASE, 14 November 1997 (1997-11-14), XP002136850 HEIDELBERG DE AC AA661040 the whole document	21-24, 30

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

**\* Special categories of cited documents :**

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

3 May 2000

Date of mailing of the international search report

22/05/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Ceder, O

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>JELSEMA ET AL.: "Phospholipid metabolism of wheat grains: Enzymes of CDP-diglyceride phospholipid biosynthetic pathway"            PROTOPLASMA,            vol. 113, 1982, pages 120-126, XP000905201            abstract</p>	1,11
A	<p>ROBINSON ET AL.: "Solubilization of microsomal associated phosphatidyl inositol synthase EC-2.7.8.11 from germinating Glycine-max cultivar cutler-71"            PLANT PHYSIOLOGY,            vol. 69, 1982, pages 146-149, XP000905186            abstract</p>	1,21
A	<p>JUSTIN ET AL.: "Compared selectivities of the phosphatidylinositol-synthase from maize coleoptiles either in microsomal membranes or after solubilization"            BIOCHIMICA ET BIOPHYSICA ACTA,            vol. 1255, no. 2, 1995, pages 161-166, XP000905179            abstract</p>	1,21